

## SPECIFICATION

## NOVEL VEGF-LIKE FACTOR

Technical Field

The present invention relates to a protein factor involved in angiogenesis in humans and falls in the field of genetic engineering.

Background Art

The process of angiogenesis, in which endothelial cells existing in the inner wall of blood vessels of animals generate new blood vessels, is triggered by transduction of a specific signal. A variety of substances are reportedly involved in this signal transduction. The most notable substance among them is the vascular endothelial growth factor (VEGF). VEGF is a protein factor which was isolated and purified, and can increase the proliferation of endothelial cells and the permeability of blood vessels (Senger, D. R. et al., Science 219: 983-985 (1983); Ferrara, N. and Henzel, W. J., Biochem. Biophys. Res. Commun. 161: 851-858 (1989)). It has been reported that the human VEGF gene contains eight exons and produces four subtypes consisting of 121, 165, 189, or 206 amino acid residues, depending on the difference in splicing, which causes different secretion patterns (Houck, K. A. et al., Mol. Endocrinol. 5: 1806-1814 (1991)). It has also been reported that there is a VEGF-specific receptor, flt-1, and that the binding of VEGF to flt-1 is important for the signal transduction (Vries, C. D. et al., Science 255: 989-991 (1992)).

Placental growth factor (PlGF) and platelet-derived growth factor (PDGF) have thus far been isolated and are factors related to VEGF. These factors are found to promote proliferation activities of vascular endothelial cells (Maglione, D. et al., Proc. Natl. Acad. Sci. USA 88: 9267-9271 (1991); Betsholtz, C. et al., Nature 320: 695-699 (1986)). In addition, VEGF-B (Olofsson, B. et al., Proc. Natl. Acad. Sci. USA 93: 2576-2581 (1996)) and VEGF-C (Lee, J. et al., Proc. Natl. Acad. Sci. USA 93: 1988-1992 (1996); Joukov, V. et al., EMBO J. 15, 290-299 (1996)) have recently been isolated.

These factors appear to constitute a family, and this may contain additional unknown factors.

It has been suggested that VEGF is involved in not only vascular formation at the developmental stage but also in the pathological neovascularization associated with diabetes, rheumatoid arthritis, retinopathy, and the growth of solid tumors. Furthermore, in addition to its vascular endothelial cell growth-promoting effects listed above, VEGF's ability to increase vascular permeability was suggested to be involved in the edema formation resulting from various causes. Also, these VEGF family factors may act on not only the blood vessels but also the blood cells and the lymphatic vessels. They may thus play a role in the differentiation and proliferation of blood cells and the formation of lymphatic vessels. Consequently, the VEGF family factors are presently drawing extraordinary attention for developing useful, novel drugs.

#### Disclosure of the Invention

An objective of the present invention is to isolate a novel protein belonging to the VEGF family and a gene encoding the protein. We searched for genes having homology to VEGF-C, which is a recently cloned VEGF family gene, against Expressed Sequence Tags (EST) and Sequence Tagged Sites (STS) in the GenBank database. As a result, we found an EST that was assumed to have homology to the C-terminal portion of VEGF-C. We then designed primers based on the sequence, and amplified and isolated the corresponding cDNA using the 5' RACE method and the 3' RACE method. The nucleotide sequence of the isolated cDNA was determined, and the deduced amino acid sequence therefrom revealed that the amino acid sequence had significant homology to that of VEGF-C. Based on the homology, we have assumed that the isolated human clone is a fourth member of the VEGF family (hereinafter designated as VEGF-D). We have also succeeded in expressing the protein encoded by the isolated human VEGF-D gene in *E. coli* cells, and have also purified and isolated it. Furthermore, we have succeeded in isolating the mouse and rat VEGF-D genes using the isolated human VEGF-D gene.

In particular, the present invention relates to a novel protein

belonging to the VEGF family and a gene encoding the protein. More specifically it relates to

- CA  
5/10/00
- (1) A protein shown by SEQ ID NO: 1 or having the amino acid sequence derived therefrom in which one or more amino acids are substituted, deleted, or added;
  - (2) A protein encoded by a DNA that hybridizes with the DNA shown by SEQ ID NO. 2;
  - (3) A DNA encoding the protein of (1);
  - CA  
5/10/00
  - (4) A DNA hybridizing with the DNA shown by SEQ ID NO: 2;
  - (5) A vector containing the DNA of (3) or (4);
  - (6) A transformant carrying the vector of (5);
  - (7) A method of producing the protein of (1) or (2), which comprises culturing the transformant of (6);
  - (8) An antibody binding to the protein of (1) or (2);
  - (9) A method of screening a compound binding to the protein of (1) or (2), which comprises a step of detecting the activity of the protein of (1) or (2) to bind to a test sample; and
  - (10) A compound binding to the protein of (1) or (2), wherein said compound has been isolated by the method of (9).

The protein of the present invention (VEGF-D) has significant homology to VEGF-C and can be considered to be a fourth factor of the VEGF family. Since the major function of VEGF is vascular formation at the developmental stage and VEGF is considered to be involved in the pathological neovascularization associated with diabetes, rheumatoid arthritis, retinopathy, and the growth of solid tumors, the protein of the present invention is thought to have similar functions.

CA  
5/10/00

A person skilled in the art could prepare functionally equivalent proteins through modifying VEGF-D of the present invention by adding, deleting, or substituting one or more of the amino acids of VEGF-D shown by SEQ ID NO: 1 using known methods. Modifications of the protein can also occur naturally in addition to the artificial modifications described above. These modified proteins are also included in the present invention. Known methods for adding, deleting, or substituting amino acids include the overlap extension polymerase chain reaction (OE-PCR) method (Gene, 1989, 77 (1): 51).

CA 5/14/00  
CA 5/14/00  
CA 5/14/00

The DNA encoding VEGF-D of the present invention, shown by SEQ ID NO: 2, is useful for isolating DNAs encoding the proteins having similar functions to VEGF-D in other organisms. For example, a person skilled in the art could routinely isolate homologs of human VEGF-D of the present invention from other organisms by allowing the DNA shown by SEQ ID NO: 2, or part thereof, as a probe, to hybridize with the DNA derived from other organisms. The DNA that hybridizes with the DNA shown by SEQ ID NO: 2 is also included in the present invention. The other organisms include mice, rats, and rabbits.

CA 5/14/00

The DNA encoding a protein that is functionally equivalent to VEGF-D usually has high homology to the DNA shown by SEQ ID NO: 2. The high homology used herein means at least 70% or higher, more preferably 80% or higher, and still more preferably 90% or higher of sequence homology.

An example of the hybridization conditions for isolating the DNA having high homology will be given below. Prehybridization is performed in ExpressHyb Solution at 68°C for 30 minutes. The probe labeled with a radioisotope is denatured at 95°C to 100°C for 2 to 5 minutes and rapidly chilled on ice. The probe is added to a new ExpressHyb Solution. The blot is transferred to the solution containing the probe and allowed to hybridize under a temperature gradient of 68°C to 55°C for 2 hours. The blot is washed four times, for 10 minute each, with a 2 x SSC solution containing 0.05% SDS at room temperature. The blot is then washed with a 0.1 x SSC solution containing 0.1% SDS at 45°C for 3 minutes. The blot is subjected to autoradiography.

An example of the hybridization conditions for isolating the DNA having very high homology will be given below. Prehybridization is performed in ExpressHyb Solution at 68°C for 30 minutes. The probe labeled with a radioisotope is denatured at 95°C to 100°C for 2 to 5 minutes and rapidly chilled on ice. The probe is added into a new ExpressHyb Solution. The blot is transferred into the solution containing the probe, and allowed to hybridize at 68°C for 1 hour. The blot was washed four times, for 10 minute each, with a 2 x SSC solution containing 0.05% SDS at room temperature. The blot was then washed with a 0.1 x SSC solution containing 0.1% SDS at 50°C for 40

minutes, during which the solution was replaced once. The blot was then subjected to autoradiography.

Note that the hybridization condition can vary depending on the length of the probe (whether it is an oligomer or a probe with more than several hundred bases), the labeling method (whether the probe is radioisotopically labeled or non-radioisotopically labeled), and the type of the target gene to be cloned. A person skilled in the art would properly select the suitable hybridization conditions. In the present invention, it is especially desirable that the condition does not allow the probe to hybridize with the DNA encoding VEGF-C.

The DNA of the present invention is also used to produce VEGF-D of the present invention as a recombinant protein. Specifically, the recombinant protein can be produced in large quantity by incorporating the DNA encoding VEGF-D (for example, the DNA shown by SEQ ID NO: 2) into a suitable expression vector, introducing the resulting vector into a host, and culturing the transformant to allow the recombinant protein to be expressed.

The vector to be used for producing the recombinant protein is not particularly restricted. However, vectors such as pGEMEX-1 (Promega) or pEF-BOS (Nucleic Acids Res. 1990, 18(17): p.5322) are preferable. / Suitable examples of the host into which the vector is introduced include E. coli cells, CHO cells, and COS cells.

The VEGF-D protein expressed by the transformant can be purified by suitably combining purification treatments such as solubilization with a homogenizer or a sonicator, extraction by various buffers, solubilization or precipitation by acid or alkali, extraction or precipitation with organic solvents, salting out by ammonium sulfate and other agents, dialysis, ultrafiltration using membrane filters, gel filtration, ion exchange chromatography, reversed-phase chromatography, counter-current distribution chromatography, high-performance liquid chromatography, isoelectric focusing, gel electrophoresis, or affinity chromatography in which antibodies or receptors are immobilized.

Once the recombinant protein is obtained, antibodies against it can be prepared using known methods. The known methods include

preparing polyclonal antibodies by immunizing rabbits, sheep, or other animals with the purified protein, and preparing monoclonal antibodies from the antibody-producing cells of immunized mice or rats. These antibodies will make it possible to quantify VEGF. Although the antibodies thus obtained can be used as they are, it will be more effective to use the humanized antibodies to reduce the immunogenicity. The methods of humanizing the antibodies include the CDR graft method and the method of directly producing a human antibody. In the CDR Graft method, the antibody gene is cloned from the monoclonal antibody-producing cells and its antigenic determinant portion is transplanted into an existing human antibody. In the method of directly producing a human antibody, a mouse whose immune system has been replaced by the human immune system is immunized, similar to ordinary monoclonal antibodies. The VEGF-D protein or its antibody thus obtained can be administered into the body by subcutaneous injection or a similar method.

A person skilled in the art could screen compounds that bind to the protein of the present invention by known methods.

For example, such compounds can be obtained by making a cDNA library on a phage vector (such as  $\lambda$ gt11 and ZAP) from the cells expected to express the protein that binds to the protein of the present invention (such as lung, small intestine, and heart cells of mammals), expressing the cDNAs on LB-agarose, fixing the expressed proteins onto a filter, preparing the purified protein of the present invention as a biotin-labeled or a fusion protein with the GST protein, and reacting this protein with the above filter. The desired compounds could then be detected by west western blotting using streptavidin or an anti-GST antibody (Skolnik, E. Y., Margolis, B., Mohammadi, M., Lowenstein, E., Fischer, R., Drepps, A., Ullrich, A., and Schlessinger, J. (1991) Cloning of P13 kinase-associated p85 utilizing a novel method for expression/cloning of target proteins for receptor tyrosine kinases, Cell 65: 83-90). Another method comprises the following steps. First, express the protein of the present invention fused with the SRF binding domain or the GAL4 binding domain in yeast cells. Second, prepare a cDNA library which expresses cDNAs fused with the transcription activation domain of VP16 or GAL4

from the cells expected to express a protein that binds to the protein of the present invention. Third, introduce the cDNA into the above yeast cells. Fourth, isolate the library-derived cDNA from the positive clones. Finally, introduce the isolated cDNA into *E. coli* to allow it to be expressed. (When a protein that binds to the protein of the present invention is expressed in yeast cells, the reporter gene is activated and the positive clone can be detected.) This method can be performed using the two-hybrid system (MATCHMAKER Two-Hybrid System, Mammalian MATCHMAKER Two-Hybrid Assay Kit, or MATCHMAKER One-Hybrid System (all by Clontech) or the HybriZAP Two-Hybrid Vector System (Stratagene) (Dalton, S. and Treisman, R. (1992) Characterization of SAP-1, a protein recruited by serum response factor to the c-fos serum response element, *Cell* 68: 597-612). Alternatively, the binding proteins can be screened by preparing a cDNA library from the cells expected to express a substance, such as a receptor, which binds to the protein of the present invention (for example, vascular endothelial cells, bone marrow cells, or lymph duct cells), introducing it into such cells as COS, detecting the binding of the protein of the present invention by itself or labeled with a radioisotope or a fluorescence, and cloning proteins that bind to the protein of the present invention (Yamasaki, K., Taga, T., Hirata, Y., Yawata, H., Kawanishi, Y., Seed, B., Taniguchi, T., Hirano, T., and Kishimoto, T. (1988) Cloning and expression of human interleukin-6 (BSF-2/IFN beta2) receptor, *Science* 241: 825-828, Fukunaga, R., Ishizaka-Ikeda, E., Seto, Y., and Nagata, S. (1990) Expression cloning of a receptor for murine granulocyte colony-stimulating factor, *Cell* 61: 341-350). Still another method comprises applying the culture supernatant or the cellular extract of the cells expected to express a protein that binds to the protein of the present invention onto an affinity column to which the protein of the present invention has been immobilized, and purifying the proteins specifically bound to the column. In addition, a DNA encoding the protein that binds to the protein of the present invention can be obtained by determining the amino acid sequence of the binding protein, synthesizing oligonucleotides based on the sequence, and screening a cDNA library with the oligonucleotides as probes.

Furthermore, compounds that bind to the protein of the present invention can be screened by contacting compounds, a natural substance bank, or a random phage peptide display library with the immobilized protein of the present invention and detecting the molecules bound to the protein. These compounds can also be screened by high throughput screening utilizing combinatorial chemistry technology (Wrighton, N. C., Farrell, F. X., Chang, R., Kashyap, A. K., Barbone, F. P., Mulcahy, L. S., Johnson, D. L., Barrett, R. W., Jolliffe, L. K., and Dower, W. J., Small peptides as potent mimetics of the protein hormone erythropoietin, Science (United States) Jul 26 1996, 273: 458-464, Verdine, G.L., The combinatorial chemistry of nature, Nature (England) Nov 7 1996, 384: 11-13, Hogan, J.C. Jr. Directed combinatorial chemistry, Nature (England) Nov 7 1996, 384: 17-19).

VEGF-D of the present invention may be used for gene therapy by introducing the VEGF-D gene into the body of the patient with the VEGF-D deficiency, or expressing the gene in the body. An anti-sense DNA of the VEGF-D gene may also be used to inhibit the expression of the gene itself, thereby suppressing the pathological neovascularization.

Among the many available methods to introduce the VEGF-D gene or its antisense DNA into the body, the retrovirus method, the liposome method, the cationic liposome method, and the adenovirus method are preferable.

In order to express these genes in the body, the genes can be incorporated into a suitable vector and introduced into the body by the retrovirus method, the liposome method, the cationic liposome method, or the adenovirus method. Although the vectors to be used are not particularly limited, such vectors as pAdexlcw and pZIPneo are preferable.

The present invention may also be applied for diagnosing disorders caused by abnormalities of the VEGF-D gene, for example, by PCR to detect an abnormality of the nucleotide sequence of the VEGF-D gene.

Furthermore, according to the present invention, the VEGF-D protein or its agonists can be used to heal wounds, promote collateral vessel formation, and aid hematopoiesis by the



hematopoietic stem cells, by taking advantage of the angiogenic effect of the VEGF-D protein. The antibodies against the VEGF-D protein or its antagonists can be used as the therapeutic agents for pathological neovascularization, lymphatic dysplasia, dyshematopoiesis, or edemas arising from various causes. The anti-VEGF-D antibodies can be used for diagnosing diseases resulting from abnormal production of VEGF-D by quantifying VEGF-D.

#### Brief Description of the Drawings

Figure 1 shows the relationship among the VEGF-D gene, the EST sequences, and the primers used for cloning.

Figure 2 compares the amino acid sequences of EST (H24828) and VEGF-C. *SEQ ID NOs: 28 and 29, respectively*

Figure 3 compares the amino acid sequences deduced from the VEGF-D gene and from the known genes of the VEGF family proteins. *SEQ ID NOs: 1 and 2, respectively*

Figure 4a shows the hydrophobicity plot of VEGF-D. Figure 4b shows the prediction of the cleavage site of the VEGF-D signal peptide.

#### Best Mode for Implementing the Invention

The following examples illustrate the present invention in detail, but are not to be construed to limit the scope of the invention.

##### Example 1. Homology search by TFASTA method

The sequence CGPNKELDENTCQCVC (SEQ ID NO: 3) was designed based on the consensus sequence found in the BR3P (Balbiani ring 3 protein) repeat at the C-terminus of VEGF-C. The entire ESTs and STS sequences in the Genbank database (as of 29 February 1996) were then searched by the TFASTA method (Person and Lipman, Proc. Natl. Acad. Sci. USA 85: 2444-2448 (1988)). The searching conditions used are shown below (Table 1).

Table 1

Sequences	392,210
Symbols	135,585,305
Word Size	2
Gap creation penalty	12.0
Gap extension penalty	4.0

As a result, an EST (Accession No. H24828) that is considered to code the consensus sequence was found. The sequence is one of the ESTs registered by The WashU-Merck EST Project, and nine out of 16 amino acid residues were identical. Further searching for UniGene by NCBI based on this sequence revealed that five registered sequences (T64149, H24780, H24633, H24828, and T64277 (as of 1 March 1996)), including the above EST, were considered to be derived from the same gene. T64277 and T64149, as well as H24828 and H24780, are the combination of the 5' sequence and the 3' sequence of the same clones, and the length of the insert in both of these clones was 0.9 kb (Fig. 1).

SEQ ID NO: 28

Translating the H24828 sequence into a protein sequence in a frame where homology is found suggested that this sequence codes 104 C-terminal amino acid residues. Comparing this amino acid sequence with the C-terminus of VEGF-C, 28 out of 104 amino acids (27%) were identical. Moreover, the amino acids that are important for maintaining the protein structure, such as cysteine and proline, were well conserved (Fig. 2). Conserved sequences are shown in a black box.

#### Example 2. cDNA cloning from a library

Primers for 5' RACE and 3' RACE (5' RACE primer: 5'-AGGGATGGGGAACCTTGAACGCTGAAT-3' (SEQ ID NO: 4), 3' RACE primer: 5'-GATCTAATCCAGCACCCCAAAACTGC-3' (SEQ ID NO: 5)) were designed (Fig. 1). A double-stranded cDNA was synthesized from human lung-derived polyA<sup>+</sup> RNA using reverse transcriptase. PCR was then performed using Marathon-Ready cDNA, Lung (Chlontech), having an adapter cDNA ligated

to both ends as a template cDNA, and using the above primer and adapter primer (AP-1 primer) as primers. The above adapter cDNA contains the regions to which the adapter primers AP-1 and AP-2 hybridize. The PCR was performed in a manner such that the system was exposed to treatment at 94°C for 1 min; five cycles of treatment at 94°C for 30 sec and at 72°C for 4 min; five cycles of treatment at 94°C for 30 sec and at 70°C for 4 min; then 25 cycles of treatment at 94°C for 20 sec and at 68°C for 4 min. (TaKaRa Ex Taq (Takara Shuzo) and the attached buffer were used as Taq polymerase instead of Advantage KlenTaq Polymerase Mix.) As a result, 1.5kb fragments were amplified at the 5' region and 0.9kb fragments at the 3' region. These fragments were cloned with the pCR-Direct Cloning System (Clontech), CR-TRAP Cloning System (GenHunter), and PT7Blue-T vector (Novagen). When the 5'-RACE fragment was cloned into the pCR-Direct vector, the fragment was amplified again using 5'-CTGGTTCGGCCCAGAACTTGAACGCTGAATCA-3' (SEQ No: 7) and 5'-CTCGCTCGCCCACTAATACGACTCACTATAGG-3' (SEQ ID NO: 8) as primers.

### Example 3. Nucleotide sequence analysis

ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit with Amplitaq DNA Polymerase FS and 377 A DNA Sequencer (ABI) were used for DNA sequencing. The primers used are the primers in the vectors (5'-AATTAACCCTCACTAAAGGG-3' (SEQ ID NO: 9), 5'-CCAGGGTTTCCAGTCACGAC-3' (SEQ ID NO: 10)), AP-2 primer (5'-ACTCACTATAGGGCTCGAGCGGC-3' (SEQ ID NO: 11)), and 10 primers in the sequence shown below (Table 2).

Table 2

SQ1 (SEQ ID NO: 12)	5'-AAGTCTGGAGACCTGCT-3'
SQ2 (SEQ ID NO: 13)	5'-CAGCAGGTCTCCAGACT-3'
SQ3 (SEQ ID NO: 14)	5'-CGCACCCAAGGAATGGA-3'
SQ4 (SEQ ID NO: 15)	5'-TGACACCTGGCCATTCCA-3'
SQ5 (SEQ ID NO: 16)	5'-CATCAGATGGTAGTTCAT-3'
SQ6 (SEQ ID NO: 17)	5'-ATGCTGAGCGAGAGTCCATA-3'
SQ7 (SEQ ID NO: 18)	5'-CACTAGGTTTGCGGCAACTT-3'
SQ8 (SEQ ID NO: 19)	5'-GCTGTTGGCAAGCACTTACA-3'
SQ9 (SEQ ID NO: 20)	5'-GATCCATCCAGATCCCTGAA-3'
SQ10 (SEQ ID NO: 21)	5'-CAGATCAGGGCTGCTTCTA-3'

Determining the nucleotide sequence of the 1.5kb fragment at the 5'-side and the 0.9kb fragment at the 3'-side revealed that the sequence of the overlapping region was identical, confirming that 5'- and 3'-side cDNAs of the desired gene were obtained. Determining the entire nucleotide sequence of the cDNA revealed that this novel gene has the full length of 2 kb and can code a protein consisting of 354 amino acid residues (SEQ ID NO: 1 and SEQ ID NO: 2). Figure 1 shows the relation between this gene and the EST sequences registered in the Genbank database. Comparing the amino acid sequence with other VEGF family proteins revealed that the amino acids that are well conserved between family proteins are also conserved in this novel gene, and therefore this gene is obviously a new member of the VEGF family (Fig. 3). In Fig. 3, HSVEGF indicates human VEGF; HSVEGF-D, HSVEGF-C, and HSVEGF-B indicate human VEGF homologues (human VEGF-D, human VEGF-C, and human VEGF-B respectively); HSPDGF-A indicates human PDGF-A; HSPDGF-B indicates human PDGF-B; and HSP1GF2 indicates human P1GF2. The conserved sequences are shown in a black box. Since VEGF-D is highly homologous to VEGF-C that was cloned as the Flt4 ligand, it was presumed to be a ligand to a Flt-4-like receptor.

Deducing the signal peptide cleavage site (Fig. 4b) by hydrophobicity plot (Fig. 4a) and the method of von Heijne (von Heijne, G, Nucleic Acids Res. 14, 4683-4690 (1986)), N-terminal 21 amino acid

residues may be cleaved as ~~signal~~ peptides, and they may also undergo additional processing like VEGF-C.

#### Example 4. Northern blot analysis

A 1kb fragment, which had been cut out by digestion with EcoRI from the 5'-fragment subcloned into pCR-Direct vector, was labeled with [ $\alpha$ -<sup>32</sup>P]dCTP and used as a probe. Labeling was performed by random priming using Ready-to Go DNA labeling beads (Pharmacia). Hybridization was performed in ExpressHyb Hybridization Solution (Clontech) by the usual method using Multiple Tissue Northern (MTN) Blot-Human, Human II, Human Fetal, and Human Cell lines (Clontech). Significant expression was observed in lung, heart, and intestine. Weak expression was observed in skeletal muscle, ovary, colon, and pancreas. The apparent molecular weight of the mRNA was 2.2 kb, and the cloned fragment seemed to be almost the full length of the gene.

#### Example 5. VEGF-D protein expression in E. coli

Two primers, 5'-TCCAGATCTTTTGC GGCAACTTTCTATGACAT-3' (SEQ ID NO: 22) and 5'-CAGGTCGACTCAAACAGGCACTAATTCAGGTAC-3' (SEQ ID NO: 23), were synthesized to amplify the region corresponding to the 89th to 181st amino acid residues of human VEGF cDNA. The thus-obtained DNA fragment was digested with restriction enzymes BglII and SalI, and ligated using ligation kit II (Takara Shuzo Co., Ltd) to plasmid pQE42 (QIAGEN), which had been digested with restriction enzymes BamHI and SalI. The resulting plasmid was introduced into E. coli SG19003[pREP4] (QIAGEN), and a plasmid, which was obtained as designed without any mutation, was selected (pQE42-BS3). Plasmid pQE42-BS3 was introduced into E. coli BL21 (Invitrogen) and cultured in 10 ml of L Broth containing 100 mg/l bicucilline (ampicillin sodium for injection, Meiji Seika Kaisha, Ltd.). 200 ml of fresh L Broth was then inoculated with the culture. After incubation at 37°C for 1.5 hours, IPTG was added to 3 mM, and the culture was further incubated at 37°C for 5 hours. After cells were harvested, a protein was purified with a Ni-NTA column following the protocol of QIAexpress TypeII kit.

#### Example 6. Expression of DHFR-VEGF-D fusion protein in E. coli

The region corresponding to the 89th to 181st amino acid residues of human VEGF cDNA was amplified with the same primers used in Example 5. The thus-obtained DNA fragment was digested with restriction enzymes BglI and SalI. The fragment was then ligated using ligation kit II (Takara Shuzo Co., Ltd.) to the plasmid pQE40 (QIAGEN), which had been digested with restriction enzymes BamHI and SalI. The resulting plasmid was introduced into E. coli SG19003[pREP4] (QIAGEN), and a plasmid, which was obtained as designed without any mutation, was selected (pQE40-BS3). Plasmid pQE40-BS3 was introduced into E. coli BL21 (Invitrogen) and cultured in 10 ml of L Broth containing 100 mg/l bicucilline (ampicillin sodium for injection, Meiji Seika Kaisha, Ltd.). 200 ml of fresh L Broth was then inoculated with the culture. After incubation at 37°C for 1.5 hours, IPTG was added to 3mM, and the culture was further incubated at 37°C for 5 hours. After cells were harvested, a DHFR-VEGF-D fusion protein was purified with a Ni-NTA column following the protocol of a QIAexpress TypeII kit.

#### Example 7. Cloning mouse VEGF-D cDNA

Two Hybond-N+ (Amersham) filters (20 cm x 22 cm) on which  $1.5 \times 10^5$  pfu of Mouse lung 5'-stretch cDNA library was transferred were prepared. Gradient hybridization from 68°C to 55°C was performed for 2 hours in ExpressHyb Hybridization Solution (Clontech) using as a probe an approximately 50 ng Pvu II fragment of human VEGF-D, which had been labeled with  $\alpha^{32}\text{P}$ -dCTP (Amersham) using Ready-To-Go DNA Labeling Beads(-dCTP) (Pharmacia). The filters were washed four times in 2 x SSC, 0.05% SDS at room temperature for 10 min, then washed in 0.1 x SSC, 0.1% SDS at 45°C for 3 min. The washed filters were exposed overnight at -80 °C using HyperFilm MP (Amersham) and intensifying paper. Positive clones were subjected to the second screening in the same manner as above to isolate a single clone. Isolated lambda DNAs were purified from the plate lysate using a QIAGEN Lambda MAX I Kit (Qiagen). Insert DNAs were cut out with EcoRI and subcloned into pUC118 EcoRI/BAP (Takara Shuzo Co., Ltd.). Its nucleotide sequence was then determined with ABI377 sequencer (Perkin

OK 5/11/00  
Elmer). The cDNA coding the full length of mouse VRGF-D was reconstructed with two of the obtained clones that overlapped each other. SEQ ID NO: 24 shows the nucleotide sequence of mouse VEGF-D cDNA and the deduced amino acid sequence therefrom.

#### Example 8. Cloning rat VEGF-D cDNA

Two Hybond-N+ (Amersham) filters (20 cm x 22 cm), on which  $1.5 \times 10^5$  pfu of Rat lung 5'-stretch cDNA library had been transferred, were prepared. Gradient hybridization from 68°C to 55°C was performed for 2 hours in ExpressH.Fyb Hybridization Solution (Clontech) using as a probe an approximately 1  $\mu$ g fragment containing 1-782 bp of the mouse VEGF-D cDNA which had been labeled with  $\alpha^{32}\text{P}$ -dCTP (Amersham) using Ready-To-Go DNA Labeling Beads (-dCTP) (Pharmacia). The filters were washed four times in 2 x SSC, 0.05% SDS at room temperature for 10 min, then washed in 0.1 x SSC, 0.1% SDS at 45°C for 3 min. The washed filters were exposed overnight at -80°C using HyperFilm MP (Amersham) and intensifying paper. Positive clones were subjected to the second screening in the same manner as above to isolate a single clone. The isolated positive clone was excised into pBluescript using E. coli SOLAR (Stratagene) and helper phage ExAssist (Stratagene), then the sequence was determined with ABI377 sequencer (Perkin Elmer). The sequence seemed to be the rat VEGF-D cDNA but did not contain the termination codon.

OK 5/11/00  
To obtain the C-terminal cDNA which had not been obtained, PCR was performed using Marathon-Ready rat kidney cDNA (Clontech) as a template and 5' primer GCTGCGAGTGTGTCTGTAAA (SEQ ID NO: 26) and 3' primer GGGTAGTGGGCAACAGTGACAGCAA (SEQ ID NO: 27) with 40 cycles of 94°C for 15 sec, 55°C for 30 sec, and 72 °C for 2 min. After the thus-obtained fragment was subcloned into pGEM-T vector (promega), the nucleotide sequence was determined with ABI377 sequencer (Perkin Elmer). The resulting clone contained the C-terminus of rat VEGF-D. Based on the results of sequencing the clone obtained by plaque hybridization and the clone obtained by PCR, the full length of the rat VEGF-D sequence was determined. SEQ ID NO: 25 shows the determined nucleotide sequence and the deduced amino acid sequence therefrom.

### Industrial Applicability

In the present invention, a novel protein (VEGF-D) having significant homology to VEGF-C and its gene have been isolated. VEGF-D appears to be involved in the pathological neovascularization associated with diabetes, rheumatoid arthritis, the growth of solid tumors, differentiation and proliferation of blood cells, formation of lymphatic vessels, and formation of edema resulting from various causes as well as the normal neovascularization at the developmental stage. The gene of the present invention can be used to diagnose disorders caused by abnormalities of the VEGF-D gene and gene therapy for the VEGF-D deficiency. The VEGF-D protein, which is obtained by expressing the gene of the present invention, can be used for healing wounds, promoting collateral vessel formation, and aiding hematopoietic stem cell proliferation. The antibodies or inhibitors against the VEGF-D protein can be used for treating angiodysplasia and lymphangiodyspasia associated with inflammation, edemas arising from various causes, dyshematopoiesis, and, as a novel anticancer agent, for treating pathological neovascularization. The VEGF-D protein and its antibodies can be useful for diagnosing diseases resulting from abnormal production of VEGF-D.



## Sequence Listing

- (1) Name or appellation of Applicant: Chugai Research Institute for Molecular Medicine, Inc.
- (2) Title of the Invention: Novel VEGF-like Factor
- (3) Reference Number: C1-802PCT
- (4) Application Number:
- (5) Filing date:
- (6) Country where the priority application was filed and the application number of the application: Japan, No. Hei 8-185216
- (7) Priority date: July 15, 1996
- (8) Number of sequences: 27

SEQ ID NO: 1

SEQUENCE LENGTH: 354

SEQUENCE TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

TISSUE TYPE: lung

SEQUENCE DESCRIPTION:

Met Tyr Arg Glu Trp Val Val Val Asn Val Phe Met Met Leu Tyr Val  
1 5 10 15  
Gln Leu Val Gln Gly Ser Ser Asn Glu His Gly Pro Val Lys Arg Ser  
20 25 30  
Ser Gln Ser Thr Leu Glu Arg Ser Glu Gln Gln Ile Arg Ala Ala Ser  
35 40 45  
Ser Leu Glu Glu Leu Leu Arg Ile Thr His Ser Glu Asp Trp Lys Leu  
50 55 60  
Trp Arg Cys Arg Leu Arg Leu Lys Ser Phe Thr Ser Met Asp Ser Arg  
65 70 75 80  
Ser Ala Ser His Arg Ser Thr Arg Phe Ala Ala Thr Phe Tyr Asp Ile  
85 90 95  
Glu Thr Leu Lys Val Ile Asp Glu Glu Trp Gln Arg Thr Gln Cys Ser  
100 105 110

Pro Arg Glu Thr Cys Val Glu Val Ala Ser Glu Leu Gly Lys Ser Thr  
 115 120 125  
 Asn Thr Phe Phe Lys Pro Pro Cys Val Asn Val Phe Arg Cys Gly Gly  
 130 135 140  
 Cys Cys Asn Glu Glu Ser Leu Ile Cys Met Asn Thr Ser Thr Ser Tyr  
 145 150 155 160  
 Ile Ser Lys Gln Leu Phe Glu Ile Ser Val Pro Leu Thr Ser Val Pro  
 165 170 175  
 Glu Leu Val Pro Val Lys Val Ala Asn His Thr Gly Cys Lys Cys Leu  
 180 185 190  
 Pro Thr Ala Pro Arg His Pro Tyr Ser Ile Ile Arg Arg Ser Ile Gln  
 195 200 205  
 Ile Pro Glu Glu Asp Arg Cys Ser His Ser Lys Lys Leu Cys Pro Ile  
 210 215 220  
 Asp Met Leu Trp Asp Ser Asn Lys Cys Lys Cys Val Leu Gln Glu Glu  
 225 230 235 240  
 Asn Pro Leu Ala Gly Thr Glu Asp His Ser His Leu Gln Glu Pro Ala  
 245 250 255  
 Leu Cys Gly Pro His Met Met Phe Asp Glu Asp Arg Cys Glu Cys Val  
 260 265 270  
 Cys Lys Thr Pro Cys Pro Lys Asp Leu Ile Gln His Pro Lys Asn Cys  
 275 280 285  
 Ser Cys Phe Glu Cys Lys Glu Ser Leu Glu Thr Cys Cys Gln Lys His  
 290 295 300  
 Lys Leu Phe His Pro Asp Thr Cys Ser Cys Glu Asp Arg Cys Pro Phe  
 305 310 315 320  
 His Thr Arg Pro Cys Ala Ser Gly Lys Thr Ala Cys Ala Lys His Cys  
 325 330 335  
 Arg Phe Pro Lys Glu Lys Arg Ala Ala Gln Gly Pro His Ser Arg Lys  
 340 345 350  
 Asn Pro

SEQ ID NO: 2

SEQUENCE LENGTH: 2004

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

TISSUE TYPE: lung

FEATURE:

NAME/KEY: CDS

LOCATION: 403..1464

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

CCAGCTTTCT GTARCTGTAA GCATTGGTGG CCACACCACC TCCTTACAAA GCAACTAGAA 60  
CCTGCGGCAT ACATTGGAGA GATTTTTTTTA ATTTTCTGGA CAYGAAGTAA ATTTAGAGTG 120  
CTTTCYAATT TCAGGTAGAA GACATGTCCA CCTTCTGATT ATTTTGGAG AACATTTTGA 180  
TTTTTTTCAT CTCTCTCTCC CCACCCCTAA GATTGTGCAA AAAAAGCGTA CCTTGCCTAA 240  
TTGAAATAAT TTCATTGGAT TTTGATCAGA ACTGATCATT TGGTTTTCTG TGTGAAGTTT 300  
TGAGGTTTCA AACTTTCCTT CTGGAGAATG CCTTTTGAAA CAATTTTCTC TAGCTGCCTG 360  
ATGTCAACTG CTTAGTAATC AGTGGATATT GAAATATTCA AA ATG TAC AGA GAG 414  
Met Tyr Arg Glu  
1  
TGG GTA GTG GTG AAT GTT TTC ATG ATG TTG TAC GTC CAG CTG GTG CAG 462  
Trp Val Val Val Asn Val Phe Met Met Leu Tyr Val Gln Leu Val Gln  
5 10 15 20  
GGC TCC AGT AAT GAA CAT GGA CCA GTG AAG CGA TCA TCT CAG TCC ACA 510  
Gly Ser Ser Asn Glu His Gly Pro Val Lys Arg Ser Ser Gln Ser Thr  
25 30 35  
TTG GAA CGA TCT GAA CAG CAG ATC AGG GCT GCT TCT AGT TTG GAG GAA 558  
Leu Glu Arg Ser Glu Gln Gln Ile Arg Ala Ala Ser Ser Leu Glu Glu  
40 45 50  
CTA CTT CGA ATT ACT CAC TCT GAG GAC TGG AAG CTG TGG AGA TGC AGG 606  
Leu Leu Arg Ile Thr His Ser Glu Asp Trp Lys Leu Trp Arg Cys Arg  
55 60 65  
CTG AGG CTC AAA AGT TTT ACC AGT ATG GAC TCT CGC TCA GCA TCC CAT 654  
Leu Arg Leu Lys Ser Phe Thr Ser Met Asp Ser Arg Ser Ala Ser His  
70 75 80  
CGG TCC ACT AGG TTT GCG GCA ACT TTC TAT GAC ATT GAA ACA CTA AAA 702  
Arg Ser Thr Arg Phe Ala Ala Thr Phe Tyr Asp Ile Glu Thr Leu Lys

85	90	95	100	
GTT ATA GAT GAA GAA TGG CAA AGA ACT CAG TGC AGC CCT AGA GAA ACG				750
Val Ile Asp Glu Glu Trp Gln Arg Thr Gln Cys Ser Pro Arg Glu Thr				
	105	110	115	
TGC GTG GAG GTG GCC AGT GAG CTG GGG AAG AGT ACC AAC ACA TTC TTC				798
Cys Val Glu Val Ala Ser Glu Leu Gly Lys Ser Thr Asn Thr Phe Phe				
	120	125	130	
AAG CCC CCT TGT GTG AAC GTG TTC CGA TGT GGT GGC TGT TGC AAT GAA				846
Lys Pro Pro Cys Val Asn Val Phe Arg Cys Gly Gly Cys Cys Asn Glu				
	135	140	145	
GAG AGC CTT ATC TGT ATG AAC ACC AGC ACC TCG TAC ATT TCC AAA CAG				894
Glu Ser Leu Ile Cys Met Asn Thr Ser Thr Ser Tyr Ile Ser Lys Gln				
	150	155	160	
CTC TTT GAG ATA TCA GTG CCT TTG ACA TCA GTA CCT GAA TTA GTG CCT				942
Leu Phe Glu Ile Ser Val Pro Leu Thr Ser Val Pro Glu Leu Val Pro				
	165	170	175	
GTT AAA GTT GCC AAT CAT ACA GGT TGT AAG TGC TTG CCA ACA GCC CCC				990
Val Lys Val Ala Asn His Thr Gly Cys Lys Cys Leu Pro Thr Ala Pro				
	185	190	195	
CGC CAT CCA TAC TCA ATT ATC AGA AGA TCC ATC CAG ATC CCT GAA GAA				1038
Arg His Pro Tyr Ser Ile Ile Arg Arg Ser Ile Gln Ile Pro Glu Glu				
	200	205	210	
GAT CGC TGT TCC CAT TCC AAG AAA CTC TGT CCT ATT GAC ATG CTA TGG				1086
Asp Arg Cys Ser His Ser Lys Lys Leu Cys Pro Ile Asp Met Leu Trp				
	215	220	225	
GAT AGC AAC AAA TGT AAA TGT GTT TTG CAG GAG GAA AAT CCA CTT GCT				1134
Asp Ser Asn Lys Cys Lys Cys Val Leu Gln Glu Glu Asn Pro Leu Ala				
	230	235	240	
GGA ACA GAA GAC CAC TCT CAT CTC CAG GAA CCA GCT CTC TGT GGG CCA				1182
Gly Thr Glu Asp His Ser His Leu Gln Glu Pro Ala Leu Cys Gly Pro				
	245	250	255	260
CAC ATG ATG TTT GAC GAA GAT CGT TGC GAG TGT GTC TGT AAA ACA CCA				1230
His Met Met Phe Asp Glu Asp Arg Cys Glu Cys Val Cys Lys Thr Pro				
	265	270	275	
TGT CCC AAA GAT CTA ATC CAG CAC CCC AAA AAC TGC AGT TGC TTT GAG				1278
Cys Pro Lys Asp Leu Ile Gln His Pro Lys Asn Cys Ser Cys Phe Glu				

280	285	290	
TGC AAA GAA AGT CTG GAG ACC TGC TGC CAG AAG CAC AAG CTA TTT CAC			1326
Cys Lys Glu Ser Leu Glu Thr Cys Cys Gln Lys His Lys Leu Phe His			
295	300	305	
CCA GAC ACC TGC AGC TGT GAG GAC AGA TGC CCC TTT CAT ACC AGA CCA			1374
Pro Asp Thr Cys Ser Cys Glu Asp Arg Cys Pro Phe His Thr Arg Pro			
310	315	320	
TGT GCA AGT GGC AAA ACA GCA TGT GCA AAG CAT TGC CGC TTT CCA AAG			1422
Cys Ala Ser Gly Lys Thr Ala Cys Ala Lys His Cys Arg Phe Pro Lys			
325	330	335	340
GAG AAA AGG GCT GCC CAG GGG CCC CAC AGC CGA AAG AAT CCT			1464
Glu Lys Arg Ala Ala Gln Gly Pro His Ser Arg Lys Asn Pro			
345	350		
TGATTCAGCG TTCCAAGTTC CCCATCCCTG TCATTTTAA CAGCATGCTG CTTTGCCAAG			1524
TTGCTGTCAC TGTTTTTTTC CCAGGTGTTA AAAAAAAT CCATTTTACA CAGCACCACA			1584
GTGAATCCAG ACCAACCTTC CATTCACACC AGCTAAGGAG TCCCTGGTTC ATTGATGGAT			1644
GTCTTCTAGC TGCAGATGCC TCTGCCACCC AAGGAATGGA GAGGAGGGGA CCCATGTAAT			1704
CCTTTTGTTT AGTTTTGTTT TTGTTTTTTC GTGAATGAGA AAGGTGTGCT GGTCATGGAA			1764
TGGCAGGTGT CATATGACTG ATTACTCAGA GCAGATGAGG AAAACTGTAG TCTCTGAGTC			1824
CTTTGCTAAT CGCAACTCTT GTGAATTATT CTGATTCTTT TTTATGCAGA ATTTGATTCG			1884
TATGATCAGT ACTGACTTTC TGATTACTGT CCAGCTTATA GTCTTCCAGT TTAATGAACT			1944
ACCATCTGAT GTTTCATATT TAAGTGATT TAAAGAAAAT AACACCATTT ATTCAAGTCT			2004

SEQ ID NO: 3

SEQUENCE LENGTH: 16

SEQUENCE TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

TISSUE TYPE: lung

SEQUENCE DESCRIPTION:

Cys Gly Pro Asn Lys Glu Leu Asp Glu Asn Thr Cys Gln Cys Val Cys

1

5

10

15

SEQ ID NO: 4

SEQUENCE LENGTH: 27

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

AGGGATGGGG AACTTGAAC GCTGAAT

27

SEQ ID NO: 5

SEQUENCE LENGTH: 27

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

GATCTAATCC AGCACCCCAA AAACCTGC

27

SEQ ID NO: 6

SEQUENCE LENGTH: 27

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

CCATCCTAAT ACGACTCACT ATAGGGC

27

SEQ ID NO: 7

SEQUENCE LENGTH: 33

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

CTGGTTCGGC CCAGAACTTG GAACGCTGAA TCA

33

SEQ ID NO: 8

SEQUENCE LENGTH: 32

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

CTCGCTCGCC CACTAATACG ACTCACTATA GG

32

SEQ ID NO: 9

SEQUENCE LENGTH: 20

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

AATTAACCCT CACTAAAGGG

20

SEQ ID NO: 10

SEQUENCE LENGTH: 22

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

CCAGGGTTTT CCCAGTCACG AC

22

SEQ ID NO: 11

SEQUENCE LENGTH: 23

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

ACTCACTATA GGGCTCGAGC GGC

23

SEQ ID NO: 12

SEQUENCE LENGTH: 17

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

AAGTCTGGAG ACCTGCT

17

SEQ ID NO: 13

SEQUENCE LENGTH: 17

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

CAGCAGGTCT CCAGACT

17

SEQ ID NO: 14

SEQUENCE LENGTH: 17

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

CGCACCCAAG GAATGGA

17

SEQ ID NO: 15

SEQUENCE LENGTH: 18

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

TGACACCTGG CCATTCCA

18



SEQ ID NO: 16

SEQUENCE LENGTH: 18

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

CATCAGATGG TAGTTCAT

18

SEQ ID NO: 17

SEQUENCE LENGTH: 20

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

ATGCTGAGCG AGAGTCCATA

20

SEQ ID NO: 18

SEQUENCE LENGTH: 20

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

CACTAGGTTT GCGGCAACTT

20

SEQ ID NO: 19

SEQUENCE LENGTH: 20

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

GCTGTTGGCA AGCACTTACA

20

SEQ ID NO: 20

SEQUENCE LENGTH: 20

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

GATCCATCCA GATCCCTGAA

20

SEQ ID NO: 21

SEQUENCE LENGTH: 19

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

CAGATCAGGG CTGCTTCTA

19

SEQ ID NO: 22

SEQUENCE LENGTH: 32

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

TCCAGATCTT TTGCGGCAAC TTTCTATGAC AT

32

SEQ ID NO: 23

SEQUENCE LENGTH: 33

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

CAGGTCGACT CAAACAGGCA CTAATTCAGG TAC

33

SEQ ID NO: 24

SEQUENCE LENGTH: 1581

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

ORIGINAL SOURCE:

ORGANISM: mouse

TISSUE TYPE: lung

FEATURE:

NAME/KEY: CDS

LOCATION: 96..1169

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

TTCCGGGCTT TGCTGGAGAA TGCCTTTTGC AACACTTTTC AGTAGCTGCC TGGAAACAAC 60  
TGCTTAGTCA TCGGTAGACA TTAAATAT TCAAATG TAT GGA GAA TGG GGA 113  
Met Tyr Gly Glu Trp Gly  
1 5  
ATG GGG AAT ATC CTC ATG ATG TTC CAT GTG TAC TTG GTG CAG GGC TTC 161  
Met Gly Asn Ile Leu Met Met Phe His Val Tyr Leu Val Gln Gly Phe  
10 15 20  
AGG AGC GAA CAT GGA CCA GTG AAG GAT TTT TCT TTT GAG CGA TCA TCC 209  
Arg Ser Glu His Gly Pro Val Lys Asp Phe Ser Phe Glu Arg Ser Ser  
25 30 35  
CGG TCC ATG TTG GAA CGA TCT GAA CAA CAG ATC CGA GCA GCT TCT AGT 257  
Arg Ser Met Leu Glu Arg Ser Glu Gln Gln Ile Arg Ala Ala Ser Ser  
40 45 50  
TTG GAG GAG TTG CTG CAA ATC GCG CAC TCT GAG GAC TGG AAG CTG TGG 305  
Leu Glu Glu Leu Leu Gln Ile Ala His Ser Glu Asp Trp Lys Leu Trp  
55 60 65 70  
CGA TGC CGG TTG AAG CTC AAA AGT CTT GCC AGT ATG GAC TCA CGC TCA 353  
Arg Cys Arg Leu Lys Leu Lys Ser Leu Ala Ser Met Asp Ser Arg Ser  
75 80 85  
GCA TCC CAT CGC TCC ACC AGA TTT GCG GCA ACT TTC TAT GAC ACT GAA 401  
Ala Ser His Arg Ser Thr Arg Phe Ala Ala Thr Phe Tyr Asp Thr Glu  
90 95 100

ACA CTA AAA GTT ATA GAT GAA GAA TGG CAG AGG ACC CAA TGC AGC CCT	449
Thr Leu Lys Val Ile Asp Glu Glu Trp Gln Arg Thr Gln Cys Ser Pro	
105 110 115	
AGA GAG ACA TGC GTA GAA GTC GCC AGT GAG CTG GGG AAG ACA ACC AAC	497
Arg Glu Thr Cys Val Glu Val Ala Ser Glu Leu Gly Lys Thr Thr Asn	
120 125 130	
ACA TTC TTC AAG CCC CCC TGT GTA AAT GTC TTC CGG TGT GGA GGC TGC	545
Thr Phe Phe Lys Pro Pro Cys Val Asn Val Phe Arg Cys Gly Gly Cys	
135 140 145 150	
TGC AAC GAA GAG GGT GTG ATG TGT ATG AAC ACA AGC ACC TCC TAC ATC	593
Cys Asn Glu Glu Gly Val Met Cys Met Asn Thr Ser Thr Ser Tyr Ile	
155 160 165	
TCC AAA CAG CTC TTT GAG ATA TCA GTG CCT CTG ACA TCA GTG CCC GAG	641
Ser Lys Gln Leu Phe Glu Ile Ser Val Pro Leu Thr Ser Val Pro Glu	
170 175 180	
TTA GTG CCT GTT AAA ATT GCC AAC CAT ACG GGT TGT AAG TGC TTG CCC	689
Leu Val Pro Val Lys Ile Ala Asn His Thr Gly Cys Lys Cys Leu Pro	
185 190 195	
ACG GGC CCC CGC CAT CCT TAC TCA ATT ATC AGA AGA TCC ATT CAG ACC	737
Thr Gly Pro Arg His Pro Tyr Ser Ile Ile Arg Arg Ser Ile Gln Thr	
200 205 210	
CCA GAA GAA GAT GAA TGT CCT CAT TCC AAG AAA CTC TGT CCT ATT GAC	785
Pro Glu Glu Asp Glu Cys Pro His Ser Lys Lys Leu Cys Pro Ile Asp	
215 220 225 230	
ATG CTG TGG GAT AAC ACC AAA TGT AAA TGT GTT TTG CAA GAC GAG ACT	833
Met Leu Trp Asp Asn Thr Lys Cys Lys Cys Val Leu Gln Asp Glu Thr	
235 240 245	
CCA CTG CCT GGG ACA GAA GAC CAC TCT TAC CTC CAG GAA CCC ACT CTC	881
Pro Leu Pro Gly Thr Glu Asp His Ser Tyr Leu Gln Glu Pro Thr Leu	
250 255 260	
TGT GGA CCG CAC ATG ACG TTT GAT GAA GAT CGC TGT GAG TGC GTC TGT	929
Cys Gly Pro His Met Thr Phe Asp Glu Asp Arg Cys Glu Cys Val Cys	
265 270 275	
AAA GCA CCA TGT CCG GGA GAT CTC ATT CAG CAC CCG GAA AAC TGC AGT	977
Lys Ala Pro Cys Pro Gly Asp Leu Ile Gln His Pro Glu Asn Cys Ser	
280 285 290	

TGC TTT GAG TGC AAA GAA AGT CTG GAG AGC TGC TGC CAA AAG CAC AAG	1025
Cys Phe Glu Cys Lys Glu Ser Leu Glu Ser Cys Cys Gln Lys His Lys	
295 300 305 310	
ATT TTT CAC CCA GAC ACC TGC AGC TGT GAG GAC AGA TGT CCT TTT CAC	1073
Ile Phe His Pro Asp Thr Cys Ser Cys Glu Asp Arg Cys Pro Phe His	
315 320 325	
ACC AGA ACA TGT GCA AGT AGA AAG CCA GCC TGT GGA AAG CAC TGG CGC	1121
Thr Arg Thr Cys Ala Ser Arg Lys Pro Ala Cys Gly Lys His Trp Arg	
330 335 340	
TTT CCA AAG GAG ACA AGG GCC CAG GGA CTC TAC AGC CAG GAG AAC CCT	1169
Phe Pro Lys Glu Thr Arg Ala Gln Gly Leu Tyr Ser Gln Glu Asn Pro	
345 350 355	
TGATTCAACT TCCTTTCAAG TCCCCCATC TCTGTCATTT TAAACAGCTC ACTGCTTTGT	1229
CAAGTTGCTG TCACTGTTGC CCACTACCCC TGCCCCCCCC CCCCCCGCC TCCAGGTGTT	1289
AGAAAAGTTG ATTTGACCTA GTGTCATGGT AAAGCCACAT TTCCATGCAA TGGCGGCTAG	1349
GTGATTCCCC AGTTCACTGA CAAATGAOTT GTAGTTTCAA ATGTCTTTGC GCCATCANCA	1409
CTCAAAAAGG AAGGGGTCTG AAGAACCCTT TGTTTGATAA ATAAAAACAG GTGCCTGAAA	1469
CAAAATATTA GGTGCCACTC GATTGGGTCC CTCGGGCTGG CCAAATTCCA AGGGCAATGC	1529
TCCTGAATTT ATTGTGCCCC TTCCTTAATG CGGAATTTCC TTTTGTTTGA TT	1581

SEQ ID NO: 25

SEQUENCE LENGTH: 1491

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

ORIGINAL SOURCE:

ORGANISM: rat

TISSUE TYPE: lung

FEATURE:

NAME/KEY: CDS

LOCATION: 270..1247

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

GCCACCTCTT GATTATTTGT GCAGCGGGAA ACTTTGAAAT AGTTTTCATC TCTTTCTCCC	60
ATACTAAGAT TGTGTGTGGC CGTGGGGGAG TCCTTGACTA ACTCAAGTCA TTTCATTGGA	120

TTTTGATTAC AACTGATCAT GTGATATTTT TTTCCATGTA AAGTTTTGGG GCTTCAAAC	180
TTGCTTCTGG AGAATGCCTT TTGCAACACT TTTCAGTAGC TGCCTGGAAA CAACTGCTTA	240
GCCATCAGTG GACATTTGAA ATATTCAAA ATG TAT GGA GAG TGG GCC GCA GTG	293
Met Tyr Gly Glu Trp Ala Ala Val	
1 5	
AAT ATT CTC ATG ATG TCC TAT GTG TAC CTG GTG CAG GGC TTC AGT ATT	341
Asn Ile Leu Met Met Ser Tyr Val Tyr Leu Val Gln Gly Phe Ser Ile	
10 15 20	
GAA CAC CGA GCA GTG AAG GAT GTT TCT CTT GAG CGA TCA TCC CGG TCT	389
Glu His Arg Ala Val Lys Asp Val Ser Leu Glu Arg Ser Ser Arg Ser	
25 30 35 40	
GTG TTG GAA CGT TCT GAA CAA CAG ATC CGC GCG GCT TCT ACT TTG GAA	437
Val Leu Glu Arg Ser Glu Gln Gln Ile Arg Ala Ala Ser Thr Leu Glu	
45 50 55	
GAG TTG CTG CAA GTC GCA CAC TCT GAG GAC TGG AAG CTG TGG CGG TGC	485
Glu Leu Leu Gln Val Ala His Ser Glu Asp Trp Lys Leu Trp Arg Cys	
60 65 70	
CGG TTG AAG CTT AAA AGT CTT GCC AAT GTG GAC TCG CGC TCA ACA TCC	533
Arg Leu Lys Leu Lys Ser Leu Ala Asn Val Asp Ser Arg Ser Thr Ser	
75 80 85	
CAT CGC TCC ACC AGA TTT GCG GCA ACT TTC TAT GAT ACT GAA ACA CTA	581
His Arg Ser Thr Arg Phe Ala Ala Thr Phe Tyr Asp Thr Glu Thr Leu	
90 95 100	
AAA GTT ATA GAT GAA GAA TGG CAG AGG ACC CAA TGC AGC CCT AGA GAG	629
Lys Val Ile Asp Glu Glu Trp Gln Arg Thr Gln Cys Ser Pro Arg Glu	
105 110 115 120	
ACA TGC GTA GAA GTC GCC AGT GAG CTG GGG AAG ACA ACC AAC ACA TTT	677
Thr Cys Val Glu Val Ala Ser Glu Leu Gly Lys Thr Thr Asn Thr Phe	
125 130 135	
TTC AAG CCC CCT TGT GTA AAT GTC TTC CGG TGT GGA GGA TGC TGC AAT	725
Phe Lys Pro Pro Cys Val Asn Val Phe Arg Cys Gly Gly Cys Cys Asn	
140 145 150	
GAA GAG AGC GTG ATG TGT ATG AAC ACA AGC ACC TCC TAC ATC TCC AAA	773
Glu Glu Ser Val Met Cys Met Asn Thr Ser Thr Ser Tyr Ile Ser Lys	
155 160 165	
CAG CTC TTT GAG ATA TCA GTG CCT CTG ACA TCA GTG CCC GAG TTA GTG	821

Gln	Leu	Phe	Glu	Ile	Ser	Val	Pro	Leu	Thr	Ser	Val	Pro	Glu	Leu	Val		
170						175					180						
CCT	GTT	AAA	ATT	GCC	AAC	CAT	ACG	GGT	TGT	AAG	TGT	TTG	CCC	ACG	GGC	869	
Pro	Val	Lys	Ile	Ala	Asn	His	Thr	Gly	Cys	Lys	Cys	Leu	Pro	Thr	Gly		
185					190					195				200			
CCC	CGG	CAT	CCT	TAT	TCA	ATT	ATC	AGA	AGA	TCC	ATT	CAG	ATC	CCA	GAA	917	
Pro	Arg	His	Pro	Tyr	Ser	Ile	Ile	Arg	Arg	Ser	Ile	Gln	Ile	Pro	Glu		
			205						210				215				
GAA	GAT	CAA	TGT	CCT	CAT	TCC	AAG	AAA	CTC	TGT	CCT	GTT	GAC	ATG	CTG	965	
Glu	Asp	Gln	Cys	Pro	His	Ser	Lys	Lys	Leu	Cys	Pro	Val	Asp	Met	Leu		
		220					225				230						
TGG	GAT	AAC	ACC	AAA	TGT	AAA	TGT	GTT	TTA	CAA	GAT	GAG	AAT	CCA	CTG	1013	
Trp	Asp	Asn	Thr	Lys	Cys	Lys	Cys	Val	Leu	Gln	Asp	Glu	Asn	Pro	Leu		
		235					240				245						
CCT	GGG	ACA	GAA	GAC	CAC	TCT	TAC	CTC	CAG	GAA	CCC	GCT	CTC	TGT	GGA	1061	
Pro	Gly	Thr	Glu	Asp	His	Ser	Tyr	Leu	Gln	Glu	Pro	Ala	Leu	Cys	Gly		
	250					255			260								
CCA	CAC	ATG	ATG	TTT	GAT	GAA	GAT	CGC	TGC	GAG	TGT	GTC	TGT	AAA	GCA	1109	
Pro	His	Met	Met	Phe	Asp	Glu	Asp	Arg	Cys	Glu	Cys	Val	Cys	Lys	Ala		
265				270					275				280				
CCA	TGT	CCT	GGA	GAT	CTC	ATT	CAG	CAC	CCG	GAA	AAC	TGC	AGT	TGC	TTT	1157	
Pro	Cys	Pro	Gly	Asp	Leu	Ile	Gln	His	Pro	Glu	Asn	Cys	Ser	Cys	Phe		
			285				290				295						
GAA	TGC	AAA	GAA	AGT	CTG	GAA	AGC	TGT	TGC	CAA	AAG	CAC	AAG	ATG	TTT	1205	
Glu	Cys	Lys	Glu	Ser	Leu	Glu	Ser	Cys	Cys	Gln	Lys	His	Lys	Met	Phe		
		300					305				310						
CAC	CCT	GAC	ACC	TGC	AGA	TCA	ATG	GTC	TTT	TCA	CTG	TCC	CCT			1247	
His	Pro	Asp	Thr	Cys	Arg	Ser	Met	Val	Phe	Ser	Leu	Ser	Pro				
	315					320					325						
TAATTTGGTT	TACTGGTGAC	ATTTAAAGGA	CATACTAACC	TGATTTATTG	GGGCTCTTTT											1307	
CTCTCAGGOC	CCAAGCACAC	TCTTAAAGGA	ACACAGACGT	TTGGCCTCTA	AGAAATACAT											1367	
GGAAGTATTA	TAGAGTGATG	ATTAAATTGT	CTTCTTGTTT	CAAACAGGGT	CTCATGATTA											1427	
CAGACCCGTA	TTGCCATGCC	TGCCGTCATG	CTATCATGAG	CGGAAAAGAA	TCACTGGCAT											1487	
TTAA																1491	

SEQ ID NO: 26

SEQUENCE LENGTH: 20

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

GCTGCGAGTG TGTCTGTAAA

20

SEQ ID NO: 27

SEQUENCE LENGTH: 25

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid, synthetic DNA

SEQUENCE DESCRIPTION:

GGGTAGTGGG CAACAGTGAC AGCAA

25